



CT Open Data Metadata Guidelines

Overview

Every dataset published on data.ct.gov is made up of two main components: data (rows, columns, etc.) and metadata (information about the data, including the description, tags, categories, etc.). Metadata helps answer the question “what is the data about?” by providing more detail about various characteristics of the data, including information about the data source, update frequency, and level of detail. It is often described as data about data. High quality metadata provides helpful context about the data’s creation, quality, and uses and is key to improving data discovery.

General Metadata Guidelines

- Consider what information is needed for a user to understand what the data is about and how to interpret it.
- Provide complete metadata to maximize public understanding of the dataset.
- Be consistent with metadata elements used across the data portal to improve data discovery.
- Avoid titles, descriptions, and other metadata elements that are confusing or difficult to interpret. Metadata should be easily understood by technical and nontechnical data users.
- Avoid the use of acronyms and use language that can be understood by a wide audience.

CT Open Data Portal Metadata Schema

The table below summarizes the metadata element on data.ct.gov. When creating a new dataset, data publishers should provide as much detail as possible in the metadata, providing the following elements where applicable.

Dataset Metadata Elements			
	Field	Definition	Permitted values
Basic Descriptive	Title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery. Avoid acronyms.	Text
	Description	What the dataset describes. Provide a longer description of the data that can be readily understood by non-technical users.	Text
	Category	The category of the data set identified by the list of possible values. If a data set can fall into multiple categories, select the one which is most significant.	Drop down menu: Business, Education, Environment and Natural Resources, Government, Health and Human Services, Housing and Development, Local Government, Public Safety, Tax and Revenue, Transportation
	Agency	The agency that collects and manages the data as the canonical source.	Text
De	Number of Rows	Number of rows in the dataset.	Auto-generated

	Row Label	What each row in the dataset represents (if applicable).	Text
	Tags/Keywords	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.	Text
	Source Link	The URL to the program area web pages	URL
	Geographic Unit	At what geographic unit is the data collected? For example, if the data is collected by address, it would be Street Address.	Drop down menu: Latitude/longitude, Street address, Intersection or street segment, Parcel (block/lot), Census block, Census block group, Census tract, Zoning district, Neighborhood, Planning District, Supervisorial District, Zip code, City, Raster, Other, Not applicable
	Temporal Coverage	This field should contain an interval of time defined by the start and end dates for which the dataset is applicable.	Date
	Related Documents/Attachments	Related documents such as technical information about a dataset, developer documentation, etc.	Attachments
Internal Management	Data Provided By	The name of the data source, for example the name of the publishing agency, organization, or individual.	Text
	Dataset Owner	The name associated with the account publishing a dataset. The name will be automatically associated with the dataset.	
	Contact Email	This address will not be displayed publicly, but inquiries submitted via the "Contact Dataset Owner" button will be routed to this email. If left blank, it will default to the dataset owner email address	Email address
Publishing Details	Update Frequency	Frequency with which dataset is updated.	Drop down menu: Not updated (historical only), As needed, Annually, Bi-annually, Quarterly, Bi-monthly, Monthly, Bi-weekly, Weekly, Daily, Hourly, Multiple times per hour
	Last Updated	Most recent date and time when the dataset was changed, updated or modified.	Auto-generated
	Public License Type	The license with which the dataset or API is published.	Drop down menu

Column Metadata

In addition to the dataset metadata (as detailed in the table above), data publishers should also provide “column metadata” (descriptions of the columns in the dataset) when appropriate. Especially when column names are not easily understood by users not familiar with the dataset, data publishers should provide short column descriptions in the metadata. Additional technical documentation can be provided as an attachment accompanying the dataset.